A SAS Data Dictionary for Achieving Analysis Dataset Compliance to FDA Guidance on Providing Submissions in Electronic Format  
Thomas Guinter, Sanofi-Aventis, Malvern, PA

ABSTRACT

The guidance documents published by CDER and CBER, in January 1999 and November 1999, respectively, have had far-reaching impact on the way clinical trials data is managed by industry and delivered to the FDA.

This presentation will provide an overview of an automated approach, using a Data Dictionary, batch programs, Macros and an AF/FSP application for achieving compliance with the guidance documents for structure of analysis datasets.

INTRODUCTION

The FDA guidance was generally received by our programmers as a “breath of fresh air.” The principles behind most of the requirements seemed to be very reasonable and represented what we considered “good business practice.” When designing CRF’s and databases for new studies and/or new compounds, we could certainly apply these principles to ensure compliance. For legacy studies, however, where CRF’s and databases were already in place, this presented significant challenges. After manually preparing two legacy compounds for submission, we decided that there must be a better way.

We developed a SAS Data Dictionary application, with SAS AF screens and batch programs running in V6.12 VAX/VMS. A combination of unique indexes and SCL behind the FSEDIT and FSVIEW screens enforces the rules in the Data Dictionary. Macros, called from the batch programs that create the analysis datasets, enforce that the datasets are created in accordance with the Data Dictionary. Utility jobs, which read the batch code as text, identify where every dataset/variable is used and store this information back into the Data Dictionary for QA reporting.

DEFINITIONS

Data Dictionary: A place where the meta data (the data about the data) is stored.

Primary Key: The combination of variables necessary to identify unique records.

APPLICATION COMPONENTS

- SAS datasets for storage of the meta data. Unique indexes are created with PROC SQL to prevent duplicate entries for primary key variables.
- SAS macros, called from batch programs that ensure the datasets are in compliance with the Data Dictionary
- SAS batch utility jobs.
- SAS AF/FSP application for entering/maintenance of the meta data

DATASETS

COMPOUNDS - The compounds for which data will be captured in the Data Dictionary. Unique index on variable COMPOUND.

<table>
<thead>
<tr>
<th>Var Name</th>
<th>Typ</th>
<th>Lbl</th>
<th>Allowable Values And/Or Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPOUND</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>FSO神秘</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO数据</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO时间</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STUDIES - The studies for which data will be captured in the Data Dictionary. Unique index on variable STUDY.

<table>
<thead>
<tr>
<th>Var Name</th>
<th>Typ</th>
<th>Lbl</th>
<th>Allowable Values And/Or Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDY</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>FSO神秘</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO数据</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO时间</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ANALYSIS - The analysis (CSR, NDA, INDAR, etc.) for which data will be captured in the Data Dictionary. Unique index on variable ANALYSIS.

<table>
<thead>
<tr>
<th>Var Name</th>
<th>Typ</th>
<th>Lbl</th>
<th>Allowable Values And/Or Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANALYSIS</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>FSO神秘</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO数据</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO时间</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CSA - The association of Compound, Study and Analysis. The starting point for a project. Composite unique index on variables COMPOUND, STUDY, ANALYSIS.

<table>
<thead>
<tr>
<th>Var Name</th>
<th>Typ</th>
<th>Lbl</th>
<th>Allowable Values And/Or Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSA</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>FSO神秘</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO数据</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSO时间</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DATASETS - The datasets that will be created for a particular Compound-Study-Analysis. Composite unique index on variables COMPOUND, STUDY, ANALYSIS, DATASET.

<table>
<thead>
<tr>
<th>Var Name</th>
<th>Typ</th>
<th>Lbl</th>
<th>Allowable Values And/Or Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPOUND</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>STUDY</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>ANALYSIS</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>DATASET</td>
<td>C</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

| FSO神秘    | Y   |     |                                    |
| FSO数据    | Y   |     |                                    |
| FSO时间    | Y   |     |                                    |
**MACROS**

**%ADINIT**

1. Setup preferred options
2. Delete the existing analysis dataset
3. Setup a libname to the data dictionary
4. Setup a global dataset macro variable (ds)

**%USAGE**

- Stores information collected by a utility job of which dataset variables are used by which programs.

**%ADDECODE**

Creates a variable containing the decode of a formatted variable.

**%ADMERGE**

Merges on common analysis variables (sex, age, race, treatment, etc.) Called from ADCREATE macro.

**%ADCREATE**

1. Validate that work dataset conforms to data dictionary
2. Call ADMERGE macro
3. Issue PROC SQL to create the permanent dataset, with variables in the sequence defined by the Data Dictionary and with labels/formats applied from the Data Dictionary
4. Create primary key inde
5. PROC CONTENTS
6. PROC PRINT

**%PRIM_KEY**

1. Determine if the requested index can be created.
   a) Validate that the dataset name and variable name(s) are correct.
   b) Validate that the variables listed will create a unique index, regardless of whether all the variables will be needed.

2. Determine if the requested index goes to the correct depth of variables necessary to guarantee uniqueness.
   a) Iteratively test each variable combination, up to one less than what is requested, to prove that the higher levels do not identify uniqueness.

3. Create the inde

**UTILITY JOBS**

**DSVCOPY**

A SAS batch job that copies/inserts observations into the Datasets-Variables dataset from/to specified compounds and/or studies.

**HOUSKEEP**

A SAS batch job, run nightly, that performs general maintenance.

1. Verify that datasets are not being edited
2. Backup all datasets (keep 7 generations)
3. Delete records flagged by FSEDIT for deletion
4. Sort the datasets
5. Renumber DS_VARS sequence
6. Recreate indexes
7. Print summary reports
8. Read programs, identifying from KEEP statements which variables are used
9. Update the Data Dictionary USAGE dataset
10. Print usage summary reports
SANOFI-SYNTHELABO BIOSTATISTICS DATA DICTIONARY

KP 1 - COMPOUNDS (CMPOUNDS)
KP 2 - STUDIES
KP 3 - ANALYSIS
KP 4 - COMPOUNDS-STUDIES-ANALYSIS (CSA)
KP 5 - DATASETS
KP 6 - VARIABLES (VARS)
KP 7 - DATASETS-VARIABLES (DSVARS)
KP 8 - FSVIEW VARIABLES (VARS)
KP 9 - FSVIEW DATASETS-VARIABLES (DSVARS)
KP 0 - FSVIEW JOIN OF VARS DSVARS

PF3 - EXIT

(*) - PRIMARY KEY VARIABLE. ONCE ENTERED/SAVED, CAN NOT BE CHANGED.
### References

- Providing Regulatory Submissions in Electronic Format - General Considerations, January 1999
- Providing Regulatory Submissions in Electronic Format - NDAs, January 1999 (CDER)
- Providing Regulatory Submissions to the Center for Biologics Evaluation and Research (CBER) in Electronic Format - Biologics Marketing Applications, November 1999

### Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Thomas Guinter  
Sanofi-Synthelabo Research  
9 Great Valley Parkway  
Malvern, PA 19355

Phone: (610) 889-6517  
Fax: (610) 889-6932  
Email: Thomas.Guinter@sanofi-synthelabo.com

---

### Sample Analysis Dataset Program

```plaintext
%adinit (centre, sr12345, wc123, csr)

data centre;
  set rd.centre (keep= add1 add2 add3 ctr_name ctr_no ctr_ty fax inv_firs inv_grad inv_name phone protid zip rename=add1+ctradd1 add2+ctradd2 add3+ctradd3 ctr_name+ctnname ctr_ty=ctnty fax+ctrfax phone+ctrphon protid=study zip+ctrsip inv_firs+invfirs inv_grad+invgrad inv_name+invname)
end=EOF;

/* create ctrty */
/***************/

%addecode (ctrty)

if ctrty=ctrty_ and ctrty=' ' then 
  %pgmerr1 (text='ctrty NOT DECODED BY FORMAT'
  ctrty=ctrty_)

/* create invgrad */
/*******************/

%addecode (invgrad)

if invgrad=invgrad_ and invgrad=' ' then 
  %pgmerr1 (text='INVGRAD NOT DECODED BY FORMAT'
  invgrad=invgrad_)

/* create center */
/*************************/

center=put (ctr_no, t____ctr.)
drop ctr_no;
/* create entry00 and entry00_ */
*******************************/
```

---

### Table: EFC1234 Study Details

<table>
<thead>
<tr>
<th>Code</th>
<th>Type</th>
<th>Description</th>
<th>Length</th>
<th>Format</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFC1234</td>
<td>CENTRE</td>
<td>Name</td>
<td>6</td>
<td>C</td>
<td>CENTER</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CNTRY00_</td>
<td>Country Code</td>
<td>15</td>
<td>C</td>
<td>CNTRY</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CNTRY00</td>
<td>Country Code</td>
<td>15</td>
<td>C</td>
<td>CNTRY</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRTY_</td>
<td>Type Code</td>
<td>15</td>
<td>C</td>
<td>CTRTY</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRTY</td>
<td>Type Code</td>
<td>15</td>
<td>C</td>
<td>CTRTY</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRFAX</td>
<td>Fax Number</td>
<td>15</td>
<td>C</td>
<td>CTRFAX</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRPHON</td>
<td>Phone Number</td>
<td>15</td>
<td>C</td>
<td>CTRPHON</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRZIP</td>
<td>Zip Code</td>
<td>15</td>
<td>C</td>
<td>CTRZIP</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRADD3</td>
<td>Address Third Line</td>
<td>15</td>
<td>C</td>
<td>CTRADD3</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRADD2</td>
<td>Address Second Line</td>
<td>15</td>
<td>C</td>
<td>CTRADD2</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRADD1</td>
<td>Address First Line</td>
<td>15</td>
<td>C</td>
<td>CTRADD1</td>
</tr>
<tr>
<td>EFC1234</td>
<td>CTRNAME</td>
<td>Name</td>
<td>15</td>
<td>C</td>
<td>CTRNAME</td>
</tr>
<tr>
<td>EFC1234</td>
<td>INVFIRS</td>
<td>First Name</td>
<td>15</td>
<td>C</td>
<td>INVFIRS</td>
</tr>
<tr>
<td>EFC1234</td>
<td>INVNAME</td>
<td>Last Name</td>
<td>15</td>
<td>C</td>
<td>INVNAME</td>
</tr>
<tr>
<td>EFC1234</td>
<td>INVGRAD</td>
<td>Grade</td>
<td>15</td>
<td>C</td>
<td>INVGRAD</td>
</tr>
</tbody>
</table>

---

### Variable Descriptions

- **CTRNAME**: Center Name (60 characters)
- **CTRTY**: Center Type (3 characters)
- **CTRFAX**: Center Fax Number (20 characters)
- **CTRPHON**: Center Phone Number (20 characters)
- **CTRZIP**: Center Zip Code (20 characters)
- **CTRADD1**: Center Address First Line (60 characters)
- **CTRADD2**: Center Address Second Line (60 characters)
- **CTRADD3**: Center Address Third Line (60 characters)
- **CTRCTRY**: Center Country (14 characters)
- **CTRCTRY_**: Center Country Code (3 characters)
- **CTRCITY**: Center City (30 characters)
- **CNTRY00**: Center Country (15 characters)
- **CNTRY00_**: Center Country Code (2 characters)

---

### Code Examples

```plaintext
/* create ctrty */
/***************/

%addecode (ctrty)

if ctrty=ctrty_ and ctrty=' ' then 
  %pgmerr1 (text='VALUE FOR CTRTY NOT DECODED BY FORMAT'
  ctrty=ctrty_)
```

```plaintext
/* create invgrad */
/*******************/

%addecode (invgrad)

if invgrad=invgrad_ and invgrad=' ' then 
  %pgmerr1 (text='VALUE FOR INVGRAD NOT DECODED BY FORMAT'
  invgrad=invgrad_)
```

---

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Thomas Guinter  
Sanofi-Synthelabo Research  
9 Great Valley Parkway  
Malvern, PA 19355

Phone: (610) 889-6517  
Fax: (610) 889-6932  
Email: Thomas.Guinter@sanofi-synthelabo.com

---

### References

- Providing Regulatory Submissions in Electronic Format - General Considerations, January 1999
- Providing Regulatory Submissions in Electronic Format - NDAs, January 1999 (CDER)
- Providing Regulatory Submissions to the Center for Biologics Evaluation and Research (CBER) in Electronic Format - Biologics Marketing Applications, November 1999